LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-JRNL-661420

Initial Global Seismic Cross-Correlation Results: Implications for Empirical Signal Detectors

D. A. Dodge, W. R. Walter

September 26, 2014

Bulletin of the Seismological Society of America

Disclaimer

Initial Global Seismic Cross-Correlation Results: Implications for Empirical Signal Detectors

D. A. Dodge[1], W. R. Walter[1]

[1]Lawrence Livermore National Laboratory
7000 East Ave, Livermore, CA 94550, USA

Sept, 2014

Douglas Dodge: dodge1@llnl.gov
William Walter: walter5@llnl.gov

Corresponding author:
Douglas Dodge
Lawrence Livermore National Laboratory
7000 East Avenue
Livermore, CA 94550
MS 046
925-423-4951

Abstract

In this work we cross-correlated waveforms in a global dataset consisting of over 310 million waveforms from nearly 3.8 million events recorded between 1970 and 2013 for two purposes: to better understand the nature of global seismicity and to evaluate correlation as a technique for automated event processing. We found that about 14.5% of the events for which we have at least one waveform correlated with at least one other event at the 0.6 or higher level. Within the geographic regions where our waveform holdings are complete or nearly complete, that fraction rose to nearly 18%. Moreover, among the events for which we had one or more seismograms recorded at distances less than 12 degrees, the fraction of correlated events was much higher, often exceeding 50%.

These results imply that global seismicity contains a large number of "repeating" events, that is, events which are sufficiently similar to each other to have correlated waveforms over the time period spanned by our dataset. These results are very encouraging for using correlation in aspects of automated event processing. It is well known that because of the strongly implied similarity of the sources of correlated signals, they can be used as empirical signal detectors (ESD), to detect, locate and identify an event using as few as one channel. Our results are very encouraging for using correlation and perhaps other forms of ESD for regional network processing and continental global processing since, for example, nearly all continental seismicity (99%) is within 12 degrees of at least one International Monitoring System station.

Introduction

It has long been known that seismic events can produce seismograms with strong similarity to previously recorded events. Quantitatively this characteristic of seismicity is often measured through waveform correlation. High correlation values between seismograms from different events imply these events have similar locations, mechanisms and other properties. Strong seismogram correlation, when it occurs, can thus be extremely useful in seismic event processing, as well as shedding light on seismic properties such as slip recurrence rates on fault patches. In this paper we attempt to better quantify how much of the Earth's seismicity is correlated and how such correlation is distributed in space and time.

Since at least the 1960's it has been known that correlation can be used as the basis for highly sensitive detectors (e.g. Anstey, 1966; Van Trees, 1968). The literature has many examples of correlation detectors applied to tightly clustered seismicity observed at local to near-regional distances; e.g. (Israelsson, 1990, Harris, 1991, Gibbons and Ringdal, 2004, 2005) to name a few. Using array-based correlation detectors, Gibbons and Ringdal (2006) demonstrated an order of magnitude reduction in the detection threshold relative to incoherent detection on a beam. These uses of correlation are so well established that at the U.S. National Data Center (USNDC), correlation detectors are routinely used for repeating sources (Junek et al., 2013). Here we treat correlation as one type of Empirical Signal Detector (ESD), a term coined by Junek et al., 2013 to refer collectively to pattern matching detectors such as correlators, subspace detectors (e.g. Harris, 2006), and matched field detectors (e.g. Harris and Kvaerna, 2010).

Correlation detectors have also been applied with some success to earthquake aftershock sequences. Large earthquake sequences are a problem for monitoring agencies because the high rate of activity can make it difficult for analysts to keep up with processing deadlines. This is due to the sheer volume of events to be processed and to the numerous false associations produced by current automated systems under conditions of high seismicity. If it is common for a significant fraction of events to be correlated, then a seismic signal processing pipeline suitably designed to use correlators to pre-group detections and prevent many false associations could far out-perform current systems during large aftershock sequences.

Harris and Dodge (2011) have used correlation in combination with subspace detectors in an automated system to track events in an aftershock sequence. They demonstrated a potential analyst workload reduction of up to 73%. Slinkard et al., (2013) applied correlation detectors to three aftershock sequences using stations from 27 to 900 km distant. They found that the percentage of bulletin events detected by correlators ranged from 30% to 92%. These examples are encouraging, but because of their local scope, cannot definitively demonstrate the potential effectiveness of correlators in a global pipeline..

Correlation detectors have also been shown to be effective over much larger regions. For example Schaff and Richards (2004, 2011) discovered that about 13% of 18,000 earthquakes in China were correlated at the CC = 0.8 level or above. At

lower correlation thresholds suitable for detection (but maybe not for location) purposes, Schaff (2009) found that two thirds of the 18,000 events were correlated. At local distances over all of northern California, Schaff and Waldhauser (2005) found that 95% of events correlate with at least one other at 4 or more stations. We find that the global average of correlated seismicity is about 18% and at short distances can rise to 50% or more. Furthermore, there is potential for higher-rank subspace detectors to improve considerably on the detection rates of pure correlators. Automated processing of 18% of world seismicity would be a significant reduction in analyst workload and the percentage of events detected by ESD is expected to grow over time. Also, a suitably designed system could mask or cancel the signals associated with all its detections. This could considerably ease the workload on the associator at times of high seismicity, resulting in fewer false associations. For these reasons it seems worthwhile to consider the use of correlators or more advanced empirical signal detectors as part of future global pipeline systems.

The present computational costs appear to be high, relative to current practice in seismology, but not by the standards of "Big Data" practitioners. For example, all channels of the IMS seismic sensors produce only a few tens of gigabytes of data per day. By comparison, in 2013 the Facebook data warehouse took in 500 terabytes per day (Miners, 2013). Implementing a system on that scale would be expensive today. However, the strong competition among vendors virtually assures that a

system designed in a few years will be able to take advantage of commodity solutions with more than enough storage and processing power.

In a future paper we will examine some of the hardware and software issues involved in scaling correlation detection to an operational capability in a global pipeline. In this paper we describe how effective correlation is expected to be; e.g. can we better quantify how much of the Earth's seismicity is correlated and how it is distributed in space, time and with what event characteristics. In this paper we attempt to answer these questions by cross correlating a large, globally distributed set of seismograms and analyzing the statistics of the resulting set of correlations. Global seismogram correlation is a very large problem and the results presented here should be considered an initial exploration of the massive results produced.

The Dataset

Lawrence Livermore National Laboratory (LLNL) operates a database of seismic events and waveforms for research on nuclear explosion monitoring and other applications. The waveforms are digital time series of ground motion recorded by seismometers installed at seismic stations. Typically, the seismometers produce output on multiple channels corresponding to different orientations and pass bands, so that often the same events are recorded on multiple channels at each station.

The LLNL database contains nearly 3.8 million events associated with more than 310 million waveforms at nearly 6,300 stations (Figure 1). The events are compiled

147 into a reconciled list from tens of individual bulletins produced by seismological
148 organizations around the world (e.g. USGS, CTBTO, ISC, numerous regional and local
149 network operators). The waveforms come from the same sources and especially
150 data collection centers such as the Incorporated Research Institutions in Seismology
151 (IRIS) Data Management Center (DMC). The figure (a) shows the completeness of
152 waveform holdings geographically. The figure was produced by gridding the Earth's
153 surface into 50km by 50km cells and, within each cell, dividing the number of events
154 for which we have at least one seismogram by the total number of events in our
155 reconciled composite global catalog for that cell. The color scale indicates the
156 completeness; with black indicating no waveforms and white indicating that for
157 every event in the cell we have at least one waveform. Although the data set has
158 global coverage, the completeness is highest in the Middle East, Eurasia, and
159 Western North America. Many of the conclusions reached in this work are based on
160 analysis of data from the regions where our coverage is 80% or greater. By
161 restricting our analysis to this subset of the data we hope to minimize biases
162 resulting from uneven distribution of waveforms in the database. The waveforms in
163 the LLNL database span a period of time greater than 60 years (b), but the earliest
164 data are for stations and channels not found later. In fact, the effective time period
165 for correlation processing is about from 1970 to the present (c).
166

167 ## Procedure

168 In order to investigate the correlation behavior of seismic signals over a wide range
169 of seismic wave types and frequencies we correlated catalog events in 8 seismic

phase windows (e.g. P, S), as well as in 15 frequency bands for each window. The bands and windows used are detailed in Tables 1 and 2. For each station-event-phase, we checked first for the existence of the phase in the AK135 travel tables. For each viable phase we chose only those bands for which there would be at least ten cycles within the nominal window length (using the band upper frequency limit). Also, to avoid duplicates when multiple branches of P- or S- existed, only a single branch was used. The windows were then arranged in time-order and trimmed as necessary to prevent overlaps. For each station-event we also processed a "whole waveform" window that extended from a few seconds before the first P to the minimum of 2000 seconds or the epicentral distance in km divided by 3.

Correlations are performed for data recorded on a common station and channel (STA-CHAN hereafter). It is impractical and unnecessary to calculate correlations for all possible event pairings per STA-CHAN. For our data set this would have required the calculation of over 10^{15} cross correlations. Rather, it is sufficient to calculate cross correlations only for those event pairs that we know to be close enough spatially that they might produce correlated seismograms. From preliminary studies we determined that it was rare for two events with correlated seismograms to have relative mislocations of more than 50 km so we chose that distance as a search radius. Although restricting the calculation of correlations only to nearby events dramatically reduces the number of correlations which must be calculated, with 3.8 million events to compare it is very important to have an efficient strategy for finding nearest neighbors. We employed a Java Spatial Index, which is the Source

193 Forge implementation of an R-Tree (Guttman, 1984). For each STA-CHAN we
194 retrieve all events recorded by that STA-CHAN, and use the R-Tree to build 'islands'
195 of events within 50 km of one-another and process all pair-wise combinations in the
196 island.
197
198 Processing of an island is shown schematically in Figure 2. An arbitrary event is
199 chosen as the starting point and the R-Tree is used to find all neighbors within 50
200 km. After measuring correlations with those neighbors, the event is removed from
201 this list and the processing is repeated with one of its neighbors. Eventually an event
202 with no neighbors is found, and the island is completely processed. The processing
203 of an event pair within an island is shown schematically in Figure 3. The waveforms
204 are retrieved (as required) and the possible windows and bands are identified. For
205 each phase and band, the seismograms are filtered and trimmed, and a signal-to-
206 noise ratio (SNR) test is performed on each window. If both windows pass the test,
207 they are correlated and if the correlation meets or exceeds 0.6, the results are
208 written to the database correlated event list. Memory and processing time
209 prevented us from writing out every correlation result, and 0.6 was chosen as an
210 interesting threshold to examine. In planned future processing using Hadoop
211 described below, we expect to be able to examine a broader suite of correlation
212 thresholds.
213
214 In developing and optimizing the algorithm discussed above, we processed several
215 subsets of the global dataset. Examination of the results revealed many instances of

correlated noise or signal artifacts. In an attempt to alleviate these problems, we tried introducing screening rules into the segment processing code. Although this achieving partial success in avoiding unwanted correlations, we decided that a more time-efficient approach was to perform the correlations without screening, and remove the invalid correlations after processing was complete. By deferring the screening, we were able to take advantage of the weeks of processing time to develop an effective algorithm. This post-processing step is discussed in detail in a later paragraph.

In all over 650 million correlations were written in about 42 days on a configuration consisting of 4 servers with 44 cores and 613 gigabytes of RAM. In addition to the correlations that were written to the database, about 700 million correlations were computed but rejected. SNR tests removed nearly 135 million windows from processing before a correlation was computed. There were nearly 678 million cases where a band was skipped because the sample rate was too low or the window was too short for the band (i.e., the window failed a simple test to prevent low time-bandwidth-product correlations). Subsequently, we re-implemented the correlation processing code using Hadoop (an open-source framework for processing large-scale data sets using commodity clusters) and achieved a speedup of nearly a factor of 20 on a test subset of events. The Hadoop implementation will enable larger and more complete investigations into correlation behavior in the future. Details of the faster Hadoop implementation are described in detail in the Addair et al. (2014) paper.

We performed post-processing of the correlation results to remove correlations due to signal artifacts and correlated noise. A significant number of seismograms used in this study contained artifacts that correlate quite well. The data from some stations were so contaminated, that tens of millions of correlations were due to artifacts. Also, our strategy of processing each phase window in multiple bands resulted in many instances of correlated noise. Examples are shown in Figure 4. Part (a) shows some of the most prevalent artifact types and part (b) shows an example of correlated noise for the case where an inappropriate filter band is used.

We were able to achieve partial success in removing signals with artifacts by computing features sensitive to each identified artifact type, and applying a threshold test for each such feature. For example, a single-point glitch can be identified using a running median filter. However, identifying optimal thresholds for each feature and establishing a prioritization of tests proved challenging. Fortunately, this kind of classification problem is well studied in computer science and a number of off-the-shelf solutions exist. We experimented with both a Support Vector Machine (Boser et al., 1992) and a random forest classifier (Breiman, 2001). We chose to use the random forest classifier because it allows inspection of the rules used in decision making.

To use the classifier, we recast our problem into two separate classification problems. In the first problem, we test our population of unfiltered signal segments

to separate valid seismograms from seismograms with non-seismic artifacts (e.g, glitches, clipping, etc.). In the second problem, the population of valid seismograms is filtered into each band used in the correlation processing, and the filtered seismogram is classified as containing a valid event seismic signal or not, based on a seismologist's assessment. These assessments of valid seismic event data are then tied to a set of computed features on which the classifier is trained. Table 3 lists the features that were computed for both filtered and unfiltered seismograms, and Table 4 lists additional features computed for the filtered segments. A number of the features in Table 3 were developed during our ad hoc attempts to remove artifacts. The remaining features attempt to describe the characteristics of the signal statistically. For example the time bandwidth product characterizes the information content, and is important in identifying signals that will produce correlations of low significance. Other features measure the way the energy is concentrated in time and frequency and the "peakedness" of the signal. The features in Table 4 are used to further characterize signals where much of the energy is concentrated in a relatively small part of the signal segment, such as a teleseismic P-arrival in a long window.

The classifier was trained using a data set of 18,300 randomly selected and filtered segments. Each segment was first presented to an analyst unfiltered, and was manually classified as being either valid or not. Those windows marked as valid were then filtered into one of the bands in which they were correlated and were presented again for judgment. The classifier was trained on the largest subset of the labeled data that preserved a 1:1 good to bad ratio (~14,000 segments).

Based on 10-fold cross validation testing, the classifier achieved about 95% precision in classification. The 10-fold validation testing consisted of forming 10 separate partitions of the data into training and validation sets, computing the precision for each partition, and averaging the results. After training and testing, the classifier was applied to all the segments referenced in the correlation results with the result that 371,209,733 correlations were retained as having been performed on valid signals.

General Characteristics of the Correlation Results

In all, 14.5% (542,405) of the 3,745,879 distinct events in our waveform table had valid correlations that met or exceeded the 0.6 acceptance threshold. Nearly 40% of the 6,266 stations produced at least one valid correlation. Figure 5 shows the distribution of the retained correlations by phase (a) and by band (b). Most of the correlations are for the whole waveform and for the S phase. Between them they account for nearly 271 million (~73%) of the correlations.

The whole-waveform window started 10 seconds before the theoretical P-wave arrival and continued to MIN (Δ_{km} / (3 km/sec), 2000 sec). Because most of the retained correlations were for relatively short event-station separations, the average length of the whole-waveform window was about 82 seconds. The effectiveness of the whole-waveform window relative to shorter windows designed to extract single phases is somewhat surprising. We initially suspected that the

correlation classifier had disproportionately removed shorter windows based on time bandwidth product values. However, examination of the removed correlations showed that the whole-waveform window was most often removed, followed by the Sn and S windows. A more likely explanation for the predominance of this window in our results is that it always exists, whereas the other windows only are computed if they are predicted by the AK135 travel time calculator for the event-station pair. Furthermore, the whole-waveform window always samples the part of the seismogram with the highest SNR whereas specific phase windows often do not.

The correlation results also are predominantly short period. Figure 5(b) shows the number of correlations as a function of filter band. The 1-2 Hz band is by far the most productive band. Most of the remaining correlations are in bands centered around or above 1 Hz. The majority of correlations were for signals recorded at local to regional distances, and at these distance ranges, (and also for teleseismic P) these are the filters one would expect to be most effective at bringing out the desired signal. Because we did not compute correlations for windows containing fewer than 10 cycles of a signal at the dominant period in any given band, there are no correlations in long-period bands at local distances or for any window other than whole-waveform. This could also contribute to most correlations being for the whole-waveform window.

Figure 6 shows the correlation counts as a function of event-station separation for long-period bands (a), mid-period bands (b) and short-period bands (c). The

331 correlations in (a) are primarily of surface waves recorded in long windows, so
332 except for the band (0.5 – 1.0Hz) there are no observations at very short distances.
333 This is a side effect of our windowing strategy as discussed previously. At mid- to
334 short-periods, the dominant feature in the plots is a drop in numbers of correlations
335 of about 3 orders of magnitude for distances greater than 8 to 10 degrees. From
336 that point to about 90 degrees, the number of correlations stays relatively constant
337 except for a bump between 35 and 51 degrees.
338
339 This behavior was surprising since our expectation was that at high frequencies,
340 attenuation of the signal (and the attendant decrease in SNR) would cause
341 decreasing correlation values with distance. To be sure that the correlations seen at
342 teleseismic distances were not dominated by misclassified artifacts we performed a
343 manual inspection of a subset of the teleseismic results. Examination of 100
344 seismogram pairs chosen randomly from the correlation results for distances of 30
345 to 90 degrees in the mid-period and short-period bands showed that in all bands
346 except one, every sample contained valid seismograms. Interestingly, nearly all
347 these teleseismic data are recorded by IMS arrays. The increase in the correlation
348 counts between about 35 and 51 degrees is a real feature. It turns out that a handful
349 of arrays are situated such that several major seismic zones fall within that distance
350 range for these arrays.
351
352 Figure 7 shows the magnitude differences (left) and the distribution of time
353 separations (right) for correlated event pairs in our results. The data are divided

into four bins based on the average magnitude of each event pair. Panel (a) shows results for Mw ≤ 2. Panel (b) shows results for 2 < Mw ≤ 4. Panel(c) shows results for 4 < Mw ≤ 6, and panel (d) shows results for 6 < Mw ≤ 8. The data were prepared by selecting all event pairs in the correlation results table for which the whole-waveform correlation exceeded 0.6 in one or more high-frequency (>0.5 Hz) bands. We are interested in understanding the detection characteristics of whole-waveform, high-frequency templates, and by restricting the data used in these plots to be high-frequency-only, we hope that the resulting statistics will be more representative of that population. The repeat interval plots were produced using these data.

Our first attempt at producing the magnitude difference distributions yielded histograms with surprisingly heavy tails. Examination of the outliers revealed that in nearly all cases, one or both of the events being compared had only a single magnitude estimate from a local or regional bulletin. Separately, we have found it to be common for such magnitude estimates to differ by a unit or more from magnitudes determined by global monitoring organizations. Therefore, we decided to remove all event pairs for which only a single magnitude estimate from a local or regional bulletin is available. This significantly reduced the number of event pairs, but there are still thousands in each magnitude range. The resulting magnitude difference histograms show that over the entire span of magnitudes in our database, events are likely to correlate well at frequencies greater than about 0.5 Hz only if their magnitudes differ by less than two units.

The histograms of repeat intervals were produced by binning the time differences of correlated events in the 4 different magnitude ranges. The most obvious feature of these plots is the abrupt ending just short of short of 20 years. This seems surprising since the time span of the waveform data is about 40 years. However, as Figure 8 shows, the LLNL waveform data can really be thought of as two distinct sets that share only a few tens of STA-CHAN between the epochs of (1970 – 1990) and (1990 – Present). At larger magnitudes, the repeat frequency decays with interval length as it must, but for Mw < 4 there is a flattening of the slope starting around 7 or 8 years. This appears to be an artifact of the way we have built our research database over many years: initially disk space limits caused us to use a short distance threshold for M<4 data collection, whereas more recently we have been collecting globally without magnitude or distance thresholds. For the largest magnitude event pairs (d) there is about an order of magnitude increase in the number of repeats in the shortest-duration bin. These are almost entirely aftershocks recorded at teleseismic distances, correlated using long windows in the 1-2 Hz.

We also calculated recurrence intervals for all the correlated event pairs found in this study, as well as in two different magnitude ranges, for time periods ranging from 1 day up to 10 years as shown in Table 5. The largest numbers of events have a 1-year or greater recurrence interval. Cumulatively we find the rate starting to approach completeness at 10 years. However as discussed in relation to Figure 8, our dataset has limited data to test very long recurrence intervals (>10 years),

because of small numbers of STA-CHAN pairs with longer operational durations. We also note a significant difference in the short time recurrence intervals; with the smaller magnitude range events have fewer short (< week) recurrence intervals than the large events. Possible explanations include the presence of significant numbers of mining and manmade repeating seismic events in the M<4 group with >1 week intervals that are absent in the large magnitude group. In addition as discussed above, our database is more complete for the large event (M>5.5) group, since disk space limits caused us to bound the collection of M<4.5 waveforms for distances beyond regional for some years. Finally we note the reconciled event catalog itself has a spatially variable completeness threshold that affects the statistics of the M<4 group.

Prevalence and Geographic Distribution of Correlated Events

The geographic distribution of correlation results as fractions of total seismicity is shown in Figure 9. To produce these plots we gridded the Earth's surface into 50km by 50km cells, and in each cell computed the ratio of correlated events to the total number of events reported in bulletins for the time period in which we have waveforms for the cell. Because we are interested in understanding the prevalence and distribution of correlated seismicity, and because the LLNL research database waveform holdings are not complete globally, we restrict most of our analysis to the region outlined by the white dashed lines. Within this region, we have waveforms for nearly all events, and therefore believe that the patterns we see in these regions are not biased by variations in data completeness.

Panel (a) shows the distribution of correlated seismicity without any restriction by band, phase, or magnitude. Globally, all or nearly all of the major seismogenic zones of the Earth are evident. The most striking features within our analysis region are the bright spots in Fennoscandia, central Asia, the Andaman Sea, and Iran. By contrast, the Mediterranean region shows a much lower fraction of correlated events. Some of these regions (e.g. Fennoscandia) have a large amount of mine seismicity which is known to correlate quite well (e.g. Tarvainen and Husebye, 1993). Panel (b) shows the distribution of correlated seismicity for events of magnitude 5 and greater. Within the analysis region, the fraction of correlated seismicity appears to be much larger on average than the distribution in (a) with most areas having a fraction greater than about 0.4. Evidently, the bright spots seen in (a) correspond to areas that have both a high density of low magnitude events and one or more stations close enough to have high SNR recordings for those events. This interpretation is supported by panel (c), which shows the fraction of events for which we have waveforms from stations within 5 degrees of the epicenters. Most of the bright spots in (a) correspond to bright spots in (c), and the Mediterranean is seen to be a region with a relatively low density of nearby stations (in our waveform database).

Evidently, correlated seismicity is not restricted geographically. But are enough events correlated to warrant making correlation detection part of routine pipeline processing? For the entire data set, about 14.5% of the events for which we have

446 one or more waveforms have mutual correlations. Within the analysis region where
447 our waveform coverage is mostly complete, the fraction increases to nearly 18%
448 and the ratio of correlated events to events reported in bulletins is nearly as high
449 (17%). Figure 10 shows the fraction of correlated seismicity as a function of source-
450 station separation in different magnitude ranges. The intent is to show how well
451 correlation detectors might perform in a system where the nearest station may be
452 several degrees from the source.
453
454 Panel (a) shows the behavior when using all possible bands and phases. For events
455 with M > 5, an astonishingly large fraction (~0.3 - 0.8) of events is correlated even at
456 very large distances. Many of these are long-period surface wave correlations, and
457 while they may not indicate the events are in close proximity, when detected at
458 multiple stations the correlated arrivals can be used to perform very accurate
459 relative locations (e.g. Cleveland and Ammon, 2013) and this could be used in
460 pipeline processing. Events with M ≤ 4 only have significant correlation fractions at
461 distances < ~10 degrees. However, for events in the range 4 < M ≤ 5 and out to
462 about 30 degrees, the correlation fraction varies from 10% to 20%. . About 10% can
463 be correlated to 70 degrees.
464
465 Panel (b) shows the behavior using only short-period bands. The correlation
466 fraction for large magnitude events averages 0.2 to 0.3 over a very large distance
467 range. This is encouraging, but should be interpreted cautiously. Nearly all these
468 correlations are for P in bands 1-2, 1-3, 2-4, and 1-5. Often these signals contain a

relatively short P-pulse followed by low-amplitude coda. For example, Figure 11 shows 80s long seismograms recorded at station KK01 for a group of 15 events correlated in the 1-2 Hz band. The correlation windows used at KK01 were about 35s long. Most of the similarity occurs within about the first 20 s. In such narrow-band, short-window cases the correlation can provide excellent relative timing between these P phases but is unlikely to indicate the causative signals are very closely located to each other. More likely, they are separated by a few tens of km. (The bulletin locations indicate a maximum separation of about 70 km.) This level of resolution may still be useful for association, or relative location based on network results but is insufficient for assignment of location based on single-station correlation, for example. Over the remaining magnitude ranges in Figure 10 (b) the behavior is similar to that of (a): The correlation fraction is large at less than ten degrees, and only the magnitude 4-5 events have a significant correlation fraction at greater distances.

Panel (c) shows the behavior in short-period bands and using a correlation threshold of 0.8. Such conditions might be required if the correlations are to be used to offload work from the associator by directly classifying new events. With these restrictions, a significant fraction of events that correlate can only be found at distances less than about 8 degrees.

Utility of Correlation Detectors for Global Seismic Monitoring

Clearly, correlation detectors (and ESDs in general) can be expected to be useful for local to regional monitoring systems. This is, after all, the domain in which many successes have been reported, and is the distance range in which this study finds the greatest fraction of correlated waveforms. In addition, our results suggest that ESDs can play an important role in a global monitoring system as well. For example, the International Monitoring System (IMS) of the Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO) will have, when complete, 50 primary and 120 auxiliary seismic monitoring stations (Brely, 2010). We estimated the fraction of continental seismicity at different distance ranges for both the primary network and for the augmented network using data from the LLNL combined bulletin (including events without waveforms). We selected a subset of events from our bulletin that were located within one of 32 seismic regions (Flynn et al., 1974) within continental areas or within an active seismic zone bordering a continent. Using this set of 5,300,239 events we calculated distances to all IMS stations. The results are listed in Table 6. For the primary network alone, 84.3% of events are within 12° of at least one station, and almost 50% are within 6°. For the full network, 99% of events are within 12° and 92.5% are within 6°. Figure 12 shows the distribution graphically. In the figure, the small circles are each centered on an IMS station and have radii of 12 degrees. Each panel shows the Earth's seismicity color-coded according to distance from the nearest station. The top shows the situation when just the primary stations are used and the bottom shows the situation using both primary and auxiliary stations. This analysis does not take into account ambient noise levels and other

513 factors that may make a station less useful. But it does suggest that a very large
514 fraction of the IMS stations may perform usefully in an ESD subsystem.
515
516 The actual design of a full-scale ESD system for a large network such as the IMS
517 would be a complex undertaking and is beyond the scope of this paper. However,
518 there are certain characteristics that we believe such a system must possess. For
519 monitoring applications, we think that although an ESD system can serve as a means
520 for lowering the detection threshold at known test sites, the system should serve
521 mainly to reduce the workload (particularly during times of high seismicity). In the
522 monitoring mission, the events of real interest are those that might be from a test
523 site or that have not been seen before and that have explosion-like characteristics. If
524 the remaining events could be masked out somehow, more resources could be
525 devoted to the events that matter. This is what we see as the principal role for an
526 ESD system: to keep on hand a pattern for every event that has ever been processed,
527 and to use those patterns to identify and screen repeat occurrences. This is a
528 computationally challenging problem because it requires the correlation of
529 terabytes to petabytes worth of templates against a real time stream. Fortunately,
530 problems of this scale are being tackled in the commercial sector and commodity
531 architectures e.g. Hadoop and Storm are being rapidly developed, and could likely
532 be adapted for this problem.
533

534 ## Discussion

In order to understand better the characteristics of global seismicity and evaluate the utility of seismic waveform correlation in automated event processing systems, we performed a very large scale global cross-correlation on a research database containing more than 300 million seismic waveforms. To understand better the dependence of waveform correlation behavior on time-bandwidth characteristics we performed the correlations in multiple time windows and frequency bands. After eliminating problematic non-seismic signal waveforms, we created a database table with about 371 million correlated seismograms. We are still examining these results in detail. In this paper, we described the most general characteristics of the results: the time, frequency, distance, and magnitude relationships between the events that showed strong correlation. In particular we are motivated by the potential to use such waveform correlation characteristics in future automated processing systems, both to lower detection thresholds and reduce the workload of human analysts.

A major potential application of seismic waveform correlation would be as part of empirical signal detectors (ESD) (e.g. correlation, subspace, matched-field, etc.). These are well known to be highly sensitive relative to power detectors. In addition, because seismic sources only produce correlated signals if the sources are very similar in location and mechanism, ESDs can detect, locate, and identify a source using as little as one channel. Because of these advantages, ESDs have been considered as components in pipeline architectures. To date, however, there have

been no large-scale deployments. The barrier to deployment is high and includes the following factors:

1. Existing pipeline architectures are very mature, and for the most part do their job very well without resort to correlation detection. Operators of these systems necessarily must be conservative about making major changes to these systems.
2. Although correlation detectors have been shown to work well in a number of regions, heretofore, it is unknown how effective they would be on a global scale.
3. Large-scale correlation processing is computationally expensive, and cannot work on the architectures currently used by pipeline operators.

We did not address the first item, but here point out that the current monitoring architecture is decades old and will eventually need to be replaced. We suggest that any redesign of a pipeline processing system should keep ESD in mind.

This paper primarily focused on the second question, global effectiveness. We found that about 14.5% of the events share at least one waveform correlation with another event (correlation coefficient >= 0.6). Within the geographic regions where our waveform holdings are complete or nearly complete, that fraction rose to nearly 18%. Moreover, among the events for which we had one or more seismograms recorded at distances less than 12 degrees, the fraction of correlated events was much higher, often exceeding 50%. We find these results to be very encouraging, with respect to point 2, since nearly all (99%) of continental seismicity is within 12 degrees of at least one IMS station.

Finally on the third point on computational expense, the landscape is changing very rapidly. During the course of this work, we became very aware of the computational complexity issues, and particularly of the impact of I/O on processing time. We ultimately re-implemented our correlation processor on the open-source Hadoop platform and found a nearly 20X speed improvement (Addair et al., 2014). The big-data analytics ecosystem of which Hadoop is a part is evolving rapidly and many businesses are processing huge amounts of data in real time using these technologies. We think this will lead to a viable architecture for processing streaming seismic data using correlation in the next few years.

Data and Resources

Most of the location and magnitude estimates used in this study can be obtained from the International Seismic Centre (ISC) (http://www.isc.ac.uk). Additional sources include the EDR catalog (http://earthquake.usgs.gov/regional/neic), the REB catalog prior to 2013 (http://www.pidc.org), the EHB catalog (ftp://ciei.colorado.edu/pub/user/engdahl/EHB), and the FINNE (http://www.seismo.helsinki.fi/bul/index.html), all of which are publicly available. Most of the waveform data were obtained through the Incorporated Research Institutes in Seismology (IRIS) Data Management Center (DMC) at www.iris.edu, the U.S. National Data Center (USNDC) at www.tt.aftac.gov, GEOSCOPE at geoscope.ipgp.jussieu.fr, IIEES at www.iiees.ac.ir, GEOFON at geofon.gfz-potsdam.de, and MEDNET at mednet.rm.ingv.it. Other data were obtained directly from networks in Azerbaijan, Georgia, Israel, Jordan, Kazakhstan, Kuwait, Oman, Saudi

603 Arabia, Turkey, and United Arab Emirates. The resources with URLs are accessed
604 and loaded into our database by automated software, and were likely last accessed
605 (prior to this study) around the beginning of May, 2013.
606

607 Acknowledgements

608 We thank Stan Ruppert and Terri Hauk for their long-term work to build and
609 maintain the LLNL research database. We thank Travis Addair for work on the
610 massive correlation processing. We thank Steve Myers and Dave Harris for
611 comments that improved the manuscript. We also thank Eric Chael, David Schaff,
612 and an anonymous reviewer for suggestions that significantly helped improve the
613 manuscript. This work was performed under the auspices of the U.S. Department of
614 Energy by Lawrence Livermore National Laboratory under contract DE-AC52-
615 07NA27344. Lawrence Livermore National Security, LLC. This is LLNL Contribution
616 LLNL-JRNL- 661420.
617

618 References

619 Addair, T. G., D.A. Dodge, W.R. Walter, S.D. Ruppert, Large-scale seismic signal
620 analysis with Hadoop, Computers & Geosciences, Volume 66, May 2014, Pages 145-
621 154, ISSN 0098-3004, http://dx.doi.org/10.1016/j.cageo.2014.01.014.
622
623 Anstey, N.A., 1966. Correlation Techniques—A Review,Can. J. Expl. Geophys., 2, 55–
624 82.

625 Boser, B. E., I. M. Guyon, and V. N. Vapnik, (1992). A training algorithm for optimal
626 margin classifiers. *In*: *COLT '92: Proceedings of the Fifth Annual Workshop on*
627 *Computational Learning Theory*. New York, NY, USA: ACM Press, pp. 144–152.

628 Breiman, Leo (2001). "Random Forests". *Machine Learning* **45** (1): 5–
629 32. doi:10.1023/A:1010933404324.
630
631 Brely, N. (2010) "The International Monitoring System", CTBTO Preparatory
632 Commission, http://ctbtcourse.files.wordpress.com/2010/10/overview-and-
633 technologies-of-ims.pdf.
634
635 Cleveland, K. M. and C. J. Ammon (2013). Precise relative earthquake location using
636 surface waves, *J. Geophys. Res., 118*, 2893-2904, Doi: 10.1002/jgrb.50146
637
638 Flinn, E. A., E. R. Engdahl, and A. R. Hill (1974). Seismic and geographical
639 regionalization, Bull. Seismol. Soc. Am. 64, 771–992.
640
641 Gibbons, S. J., and F. Ringdal (2004). A waveform correlation procedure for detecting
642 decoupled chemical explosions, NORSAR Scientific Report: Semiannual Technical
643 Summary No. 2–2004, NORSAR, Kjeller, Norway, 41–50.
644
645 Gibbons, S. J., and F. Ringdal (2005). The detection of rockbursts at the Barentsburg
646 coal mine, Spitsbergen, using waveform correlation on SPITS array data, NORSAR
647 Scientific Report: Semiannual Technical Summary No. 1–2005, NORSAR, Kjeller,
648 Norway, 35–48.
649
650 Gibbons, S., and F. Ringdal (2006). The detection of low magnitude seismic events
651 using array-based waveform correlation, Geophys. J. Int. 165, 149–166.
652
653 Guttman, A. (1984). "R-Trees: A Dynamic Index Structure for Spatial
654 Searching". *Proceedings of the 1984 ACM SIGMOD international conference on*
655 *Management of data - SIGMOD '84*. p. 47.doi:10.1145/602259.602266.
656
657 Harris, D. B., 1991. A waveform correlation method for identifying quarry
658 explosions, Bull. Seismol. Soc. Am. 80, no. 6, 2177-2193.
659
660 Harris, D. (2006), Subspace detectors: Theory, Lawrence Livermore Natl. Lab. Rep.
661 UCRL-TR-222758, 46 pp., Lawrence Livermore Natl. Lab., Livermore, Calif.
662
663 Harris, D.B., and T. Kvaerna, 2010, Superresolution with seismic arrays using
664 empirical matched field processing: Geophysical Journal International, v. 182, p.
665 1455-1477.
666
667 Harris, D. and D. Dodge (2011). An autonomous system for grouping events in a
668 developing aftershock sequence, Bull. Seism. Soc. Am. 101, 763-774,
669 doi:10.1785/0120100103.
670
671 Israelsson, H., 1990. Correlation of waveforms from closely spaced regional
672 events, *Bull. Seism. soc. Am.*, **80**(6), 2177–2193.
673

Junek, W. N., VanDeMark, T. F., Saults, T. R., Harris, D. B., Dodge, D. A., Matlagh, S., Ichinose, G. A., Poffenberger, A., and R. C. Kemerait, 2013. "Integration of Empirical Signal Detectors into the Detection and Feature Extraction Application at the United States National Data Center", poster, CTBTO Science and Technology Meeting, Vienna, Austria, 17-21 June, 2013.

Miners, Z. (2013). "Facebook's big data plans include warehouses, faster analytics", Computerworld, April 30, 2013.

Schaff, D. P., and P. G. Richards (2004). Repeating seismic events in China, Science 303, 1176-1178.

Schaff, D. P. and F. Waldhauser (2005). Waveform cross-correlation-based differential travel-time measurements at the Northern California seismic network Bull. Seismol. Soc. Am., 95(6), 2446-2461.

Schaff, D. P. (2009), Broad-scale applicability of correlation detectors to China seismicity, Geophys. Res. Lett., 36, L11301, doi:10.1029/2009GL038179.

Schaff, D. P., and P. G. Richards (2011). On finding and using repeating seismic events in and near China. J. Geophys. Res. 116, doi:10.1029/2010JB007895.

Slinkard, M. E., Carr, D. B., and C. J. Young, 2013. Applying Waveform Correlation to Three Aftershock Sequences, *Bull. Seism. Soc. Am.*, 103(2A) 675-693; doi:10.1785/0120120058.

Tarvainen, M., and E. S. Husebye (1993). Spatial and Temporal Patterns of the Fennoscandian Seismicity – an Exercise in Explosion Monitoring, Geophysica, 29 (1-2) 1-19.

Van Trees, H. L. (1968), *Detection, Estimation and Modulation Theory*, vol. 1, John Wiley and Sons, New York.

Author Addresses

Douglas Dodge
Lawrence Livermore National Laboratory
7000 East Avenue
Livermore, CA 94550
MS 046
925-423-4951
dodge1@llnl.gov

William Walter
Lawrence Livermore National Laboratory
7000 East Avenue
Livermore, CA 94550
MS 046
925-423-8777
walter5@llnl.gov

Tables

Windows used for correlation processing

PHASE	NOMINAL WINDOW LENGTH (s)	PRE-WIN SECONDS	MIN Δ^0	MAX Δ^0	MAX DEPTH
Lg	50	10	1.46	15	35
P	30	5	0	90	700
PcP	50	5	26	60	700
Pg	30	10	0	1.5	35
Pn	15	7	1.5	10	35
S	30	10	0	90	700
Sn	30	10	1.46	15	35
Whole	2000	5	0	90	700

Table 1 shows the phases for which correlations could be computed. In order for the phase to be used at a specific event-station, the event had to fall within the depth range specified by (MIN DEPTH, MAX DEPTH) and the distance to the station had to be within (MIN DELTA, MAX DELTA). The window starting positions were calculated using AK135 and extended from PRE-WIN SECONDS before the predicted arrival for NOMINAL WIN LENGTH seconds. In a case where a window would extend into another predicted phase, the window was truncated at the predicted onset of the following phase. For the phase 'Whole' the nominal window length was calculated as MIN(nominal, DELTA (km) / 3).

737 Frequency bands used for correlation processing

LOW CORNER (Hz)	HIGH CORNER (Hz)
0.025	0.05
0.05	0.1
0.1	0.2
0.5	1
1	2
2	4
4	8
0.02	0.1
0.5	5
0.75	3
1	3
1	5
2	6
3	9
4	16

738 Table 2 shows the frequency bands for which correlations might be computed. The
739 bands were chosen so that for any phase and distance there would be at least one
740 band optimum for the signal. For each window pair to be processed only those
741 bands supported by the seismogram sample rate and containing a minimum of 10
742 cycles at the band center were used.

743

744

745 Features for all segments

Feature Name	Feature Description
SNR	Max of Signal to Noise ratios at P-arrival computed using both short and long windows
Kurtosis	Unbiased estimator of population excess Kurtosis
T0	Time centroid of signal
VART	Time Variance
VARF	Frequency Variance
TBP	Time Bandwidth Product
EXTR	(mean – median) / range
NGLITCH	Number of single-point glitches
DROPOUT_FRAC	Number of discrete intervals in the trace with N or more consecutive samples having the same value.
DROPOUT_IMP	For the drop out fraction, the ratio between the (dropout mean – signal mean) and the signal range.
DISTINCT_VAL_RATIO	The number of distinct values divided by the total values (This is inversely related to the quantization error.)
NUM_DISC	Number of places where there was a sudden persistent change in the signal mean
DISC_AVG_VALUE	Average value of the discontinuities
DISC_MAX_VALUE	Max value of the discontinuities
DISC_AVG_KURTOSIS	Average kurtosis of the discontinuities
DISC_MAX_KURTOSIS	Max kurtosis of the discontinuities

746 Table 3 lists the features that are computed for both raw and filtered segments. The
747 first six features are statistical descriptions of the signal, while the remaining
748 features are used to characterize certain artifacts.
749

750 Additional features for filtered segments

Feature Name	Feature Description
Non_centrality	This is essentially a T-statistic comparing the signal window center to the energy centroid
Packet_end	The time from the window start where 90% of the energy in the envelope has been seen
Packet_centroid	Time centroid of packet
Packet_TBP	Time Bandwidth of packet
Packet_sigma	Square root of the packet time variance

751 Table 4 lists additional features computed only on filtered segments.

752

Recurrence Interval Statistics

Magnitudes		< 1 day	1 day – 1 week	1 week – 1 month	1 month – 1 year	1 – 10 year	10 – 20 year
All	Number	921435	2098473	3708529	13499236	25003447	4422321
	Percent	1.9%	4.2%	7.5%	27.2%	50.3%	8.9%
	Cumulative	1.9%	6.1%	13.6%	40.8%	91.1%	100%
M ≤ 4	Number	708383	1872547	3498688	12924089	24075716	4374775
	Percent	1.5%	3.9%	7.4%	27.2%	50.7%	9.2%
	Cumulative	1.5%	5.4%	12.8%	40%	90.7%	99.9%
M > 5.5	Number	15488	12484	10143	31123	78750	10081
	Percent	9.8%	7.9%	6.4%	19.6%	49.6%	6.3%
	Cumulative	9.8%	17.7%	24.1%	43.7%	93.3%	99.6%

Table 5 lists recurrence interval statistics for all correlated event pairs found in this study. The table is divided into three sections. The top section labeled “All” contains the statistics for all event pairs without regard to event magnitude. The middle section contains the statistics for event pairs with average magnitude ≤ 4 and the bottom section contains statistics for event pair with average magnitude > 5.5.

IMS Station Coverage of Continental Seismicity

	$\Delta^o \leq 1$	$\Delta^o \leq 2$	$\Delta^o \leq 4$	$\Delta^o \leq 6$	$\Delta^o \leq 8$	$\Delta^o \leq 10$	$\Delta^o \leq 12$
Primary	**6.2%**	**15%**	**35.9%**	**49.1%**	**63.1%**	**78.5%**	**84.3%**
All	**16.9%**	**37.7%**	**78.8%**	**92.5%**	**98.1%**	**98.8%**	**99%**

Table 6 shows the fractions of continental seismicity within distance ranges from 1° to 12° of at least 1 IMS station. The row labeled "Primary" shows percentages when only primary network stations are considered, and the row labeled "All" shows percentages for the entire network.

Figures

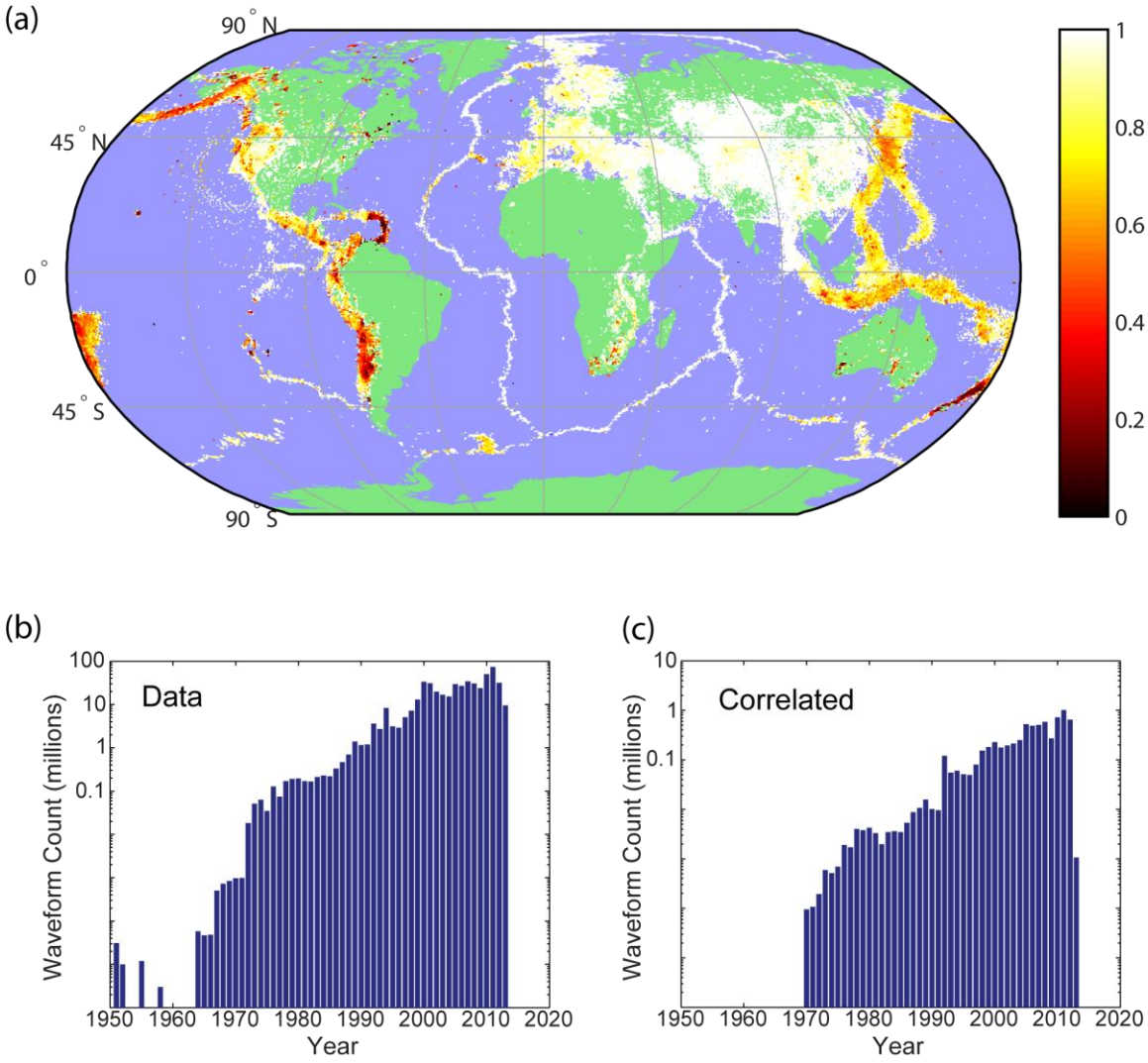
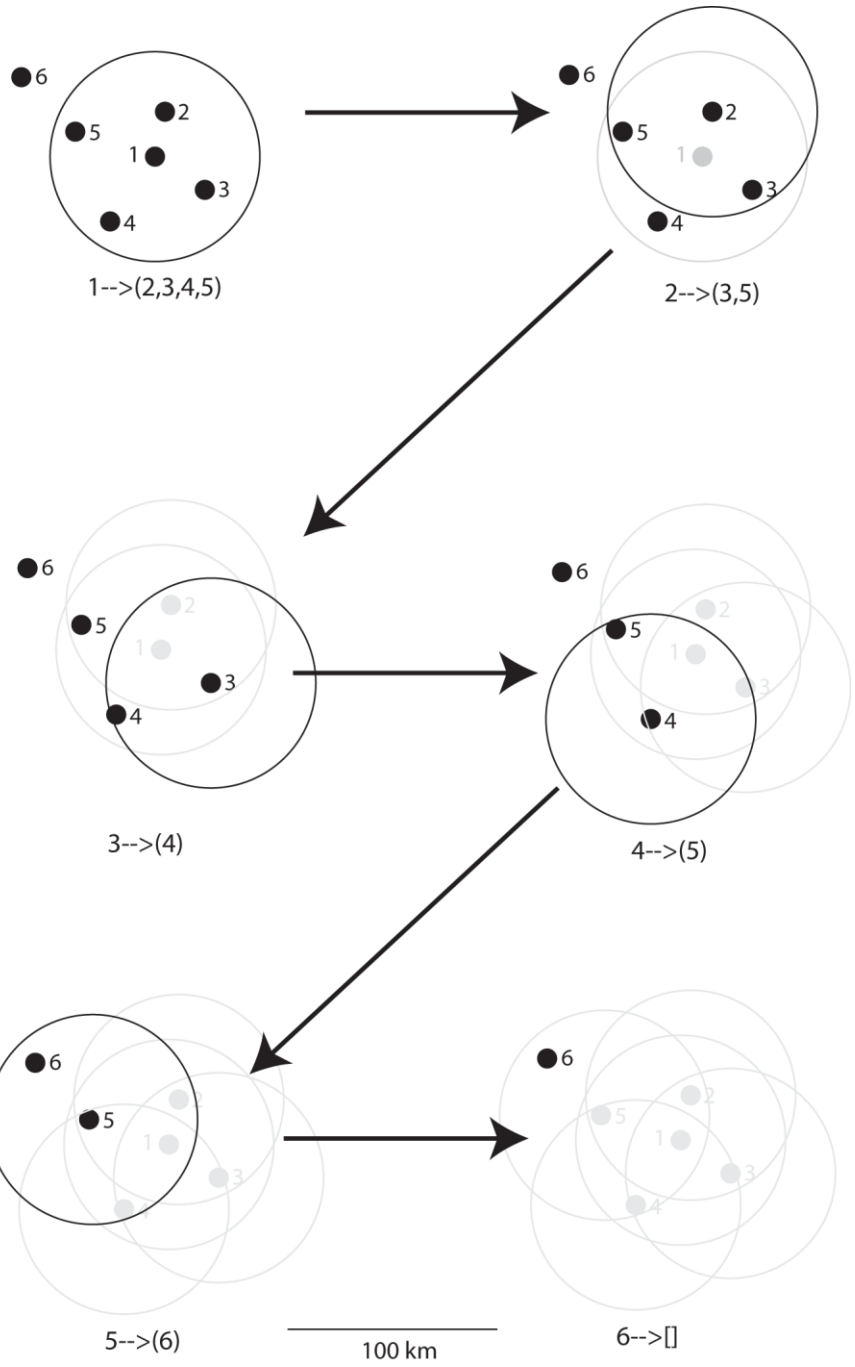


Figure 1 (a) shows the waveform completeness (number of events with waveforms per cell divided by the total number of events in the cell during the bounding epoch of the waveforms). Color is proportional to completeness with black lowest and white highest. Note that although the data set has global coverage, the completeness is highest in the Middle East, Eurasia, Fennoscandia, and Western North America. Panel (b) shows waveform segment counts by year and panel (c) shows the segment counts by year for waveform segments that eventually were found to correlate with another.



778

779 Figure 2 shows (schematically) the processing of an "island". The traversal strategy
780 minimizes I/O and computations by requiring each waveform to be read only once
781 and correlated only once with neighbors within 50 km. At each stage an R-tree is
782 used to rapidly determine candidates. At the start, events 2-5 have been found to be
783 within 50 km of (1). Waveforms for all five are loaded and (1) is processed against
784 the others for all phases and bands. At this point, all data for (1) is removed from
785 memory and the focus shifts to (2). Processing of the island continues until all
786 events have been processed.

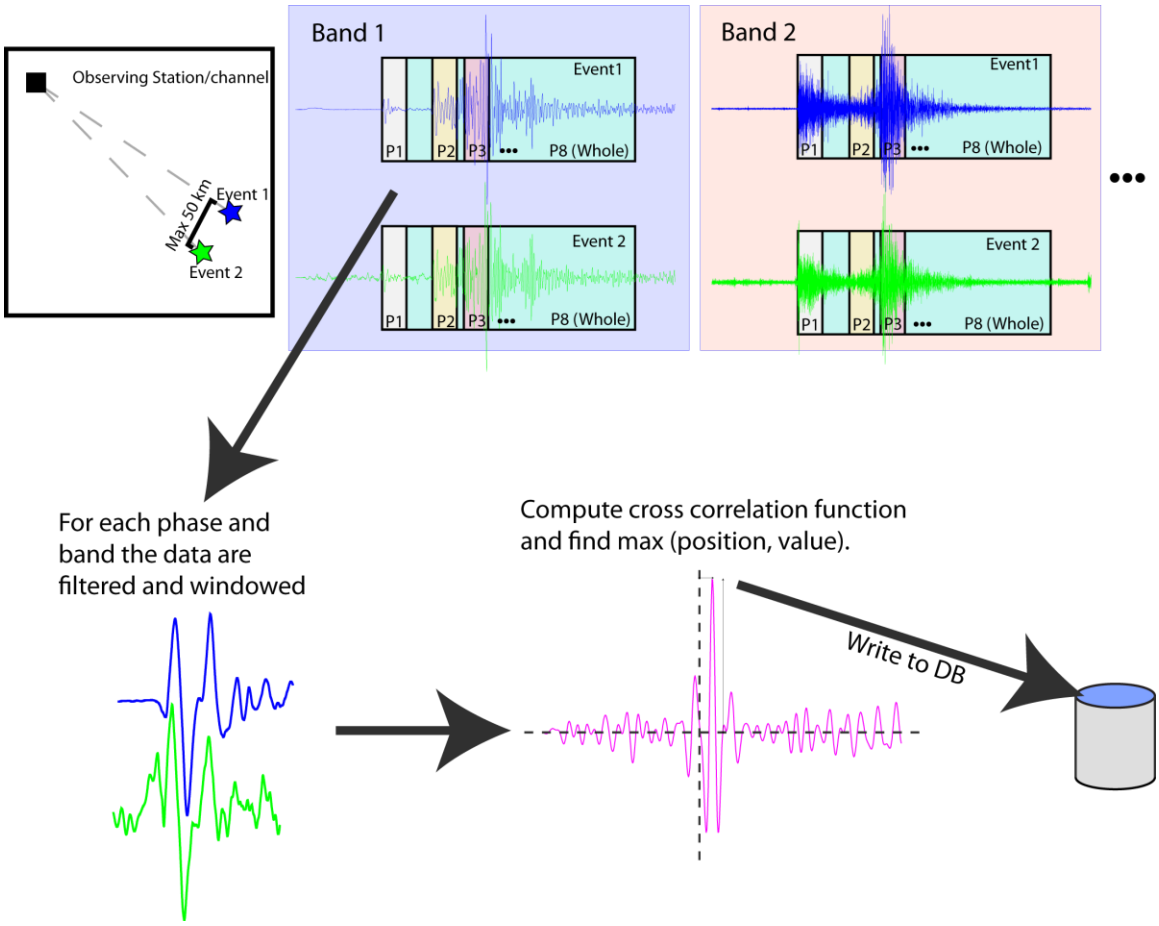


Figure 3 is a schematic illustration of the processing applied to a single channel for a pair of events observed by a single station. The graphic in the upper left shows the geometry of the station and events to be processed. The graphics labeled "Band 1" and "Band 2" show the seismogram pair filtered into two different bands, and indicate (schematically) the windows for which correlations will be computed. For each window pair, the cross correlation function is computed and the max and its associated shift are recorded in the database. This is indicated schematically in the lower part of the figure.

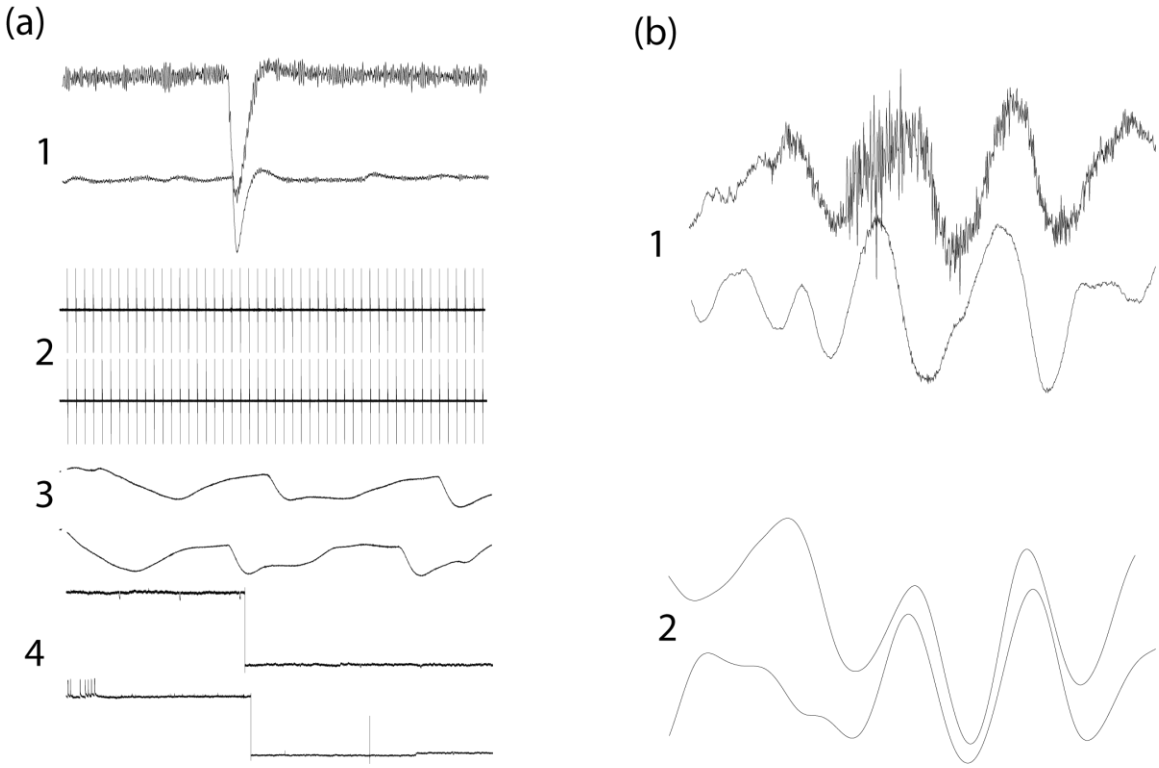


Figure 4 shows examples of common artifacts that correlate well and that were removed in a post-processing step. (a1) is an apparent calibration pulse. (a2) is a comb function due to some kind of electrical malfunction. (a3) is an unidentified artifact (perhaps sensor tilting?) that is surprisingly common on some STA-CHAN. (a4) is a step function probably due to an electrical malfunction. (b) is an example of an artifact caused by filtering a signal into a narrow band that contains noise and with the intended signal well outside the band. The top shows the raw traces with a high frequency seismogram riding on low frequency noise. After filtering into the band containing the noise, the intended signal is gone and only the narrow band noise is left. The filtered signal will correlate quite well, but the result has no seismological significance.

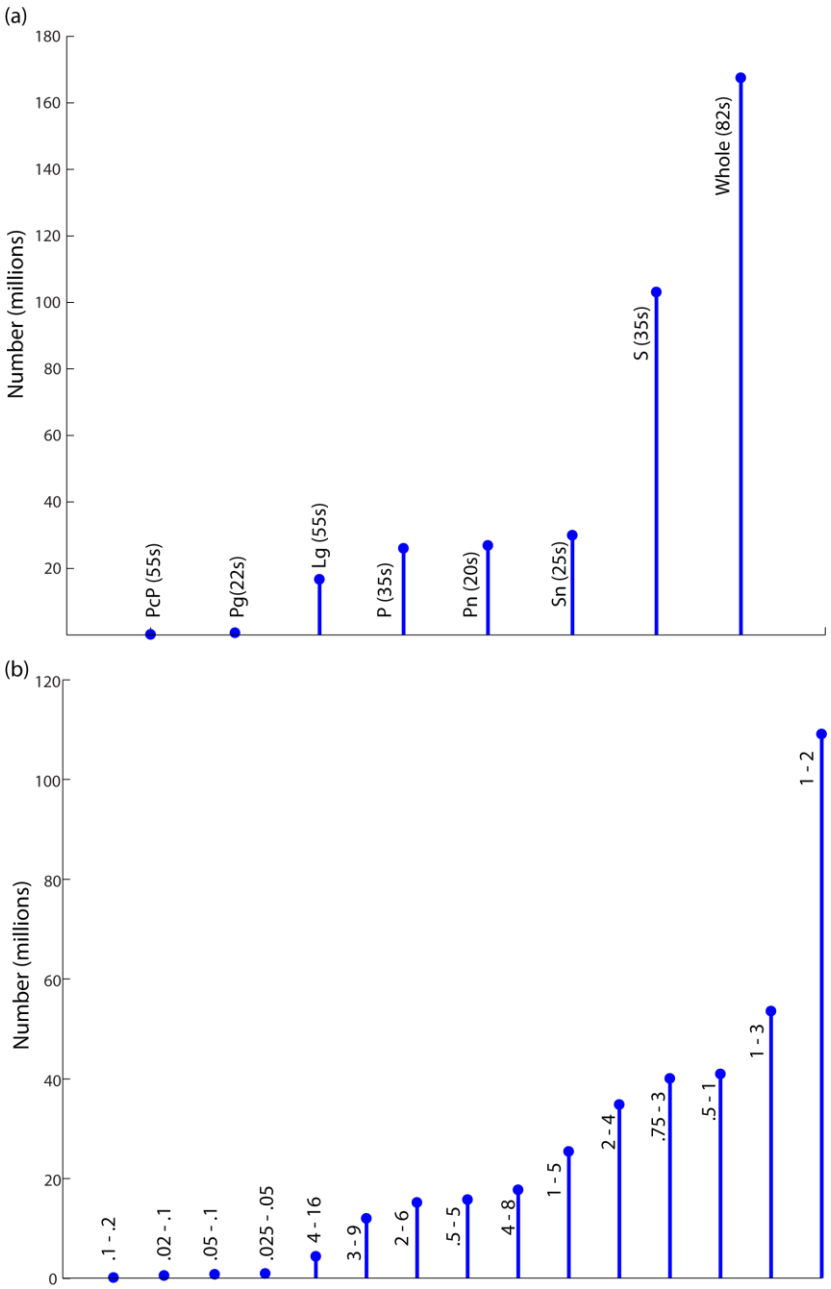
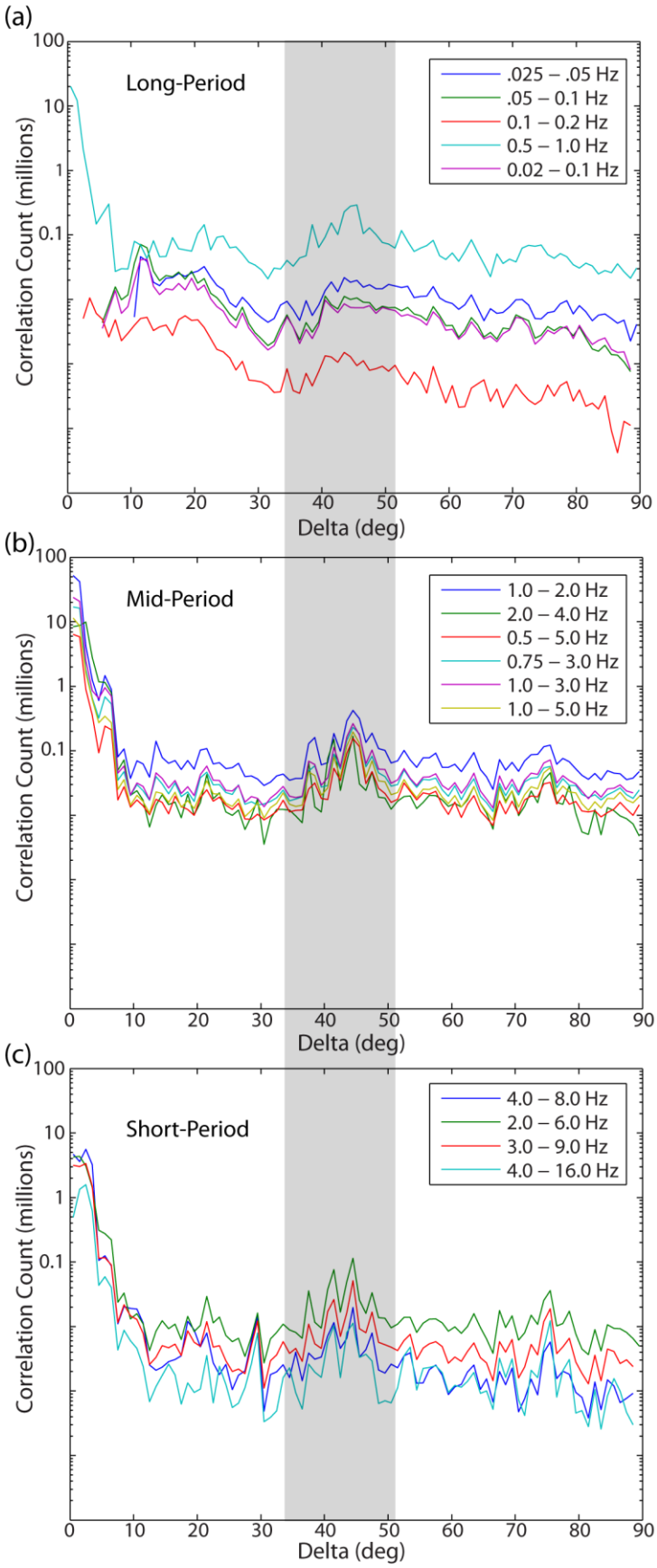


Figure 5 shows the overall distribution of correlations by phase (a) and by frequency band (b). In part (a) the labels on each "stick" indicate the phase and the average window length. For all windows except "Whole" the length was predetermined but subject to the constraint that the correlation window could not run into the next phase. The length of the "Whole" window was determined based on the source-receiver distance. Although this window could be as long as 2000s, because most of the retained correlations are for relatively short distances, the average length for this phase is only 82s. Part (b) shows the number of retained correlations as a function of filter band. The vast majority are in short-period bands.


(a)
Long-Period
.025 – .05 Hz
.05 – 0.1 Hz
0.1 – 0.2 Hz
0.5 – 1.0 Hz
0.02 – 0.1 Hz
Correlation Count (millions)
Delta (deg)
(b)
Mid-Period
1.0 – 2.0 Hz
2.0 – 4.0 Hz
0.5 – 5.0 Hz
0.75 – 3.0 Hz
1.0 – 3.0 Hz
1.0 – 5.0 Hz
Correlation Count (millions)
Delta (deg)
(c)
Short-Period
4.0 – 8.0 Hz
2.0 – 6.0 Hz
3.0 – 9.0 Hz
4.0 – 16.0 Hz
Correlation Count (millions)
Delta (deg)

820

Figure 6 shows the correlation counts as a function of event-station separation for long period bands (a), mid period bands (b) and short period bands (c). At mid to long periods the dominant feature in the plots is a drop of about 3 orders of magnitude for distances greater than 8 to 10 degrees. From that point out to about 90 degrees the number of correlations stays relatively constant except for a bump between 35 and 51 degrees.

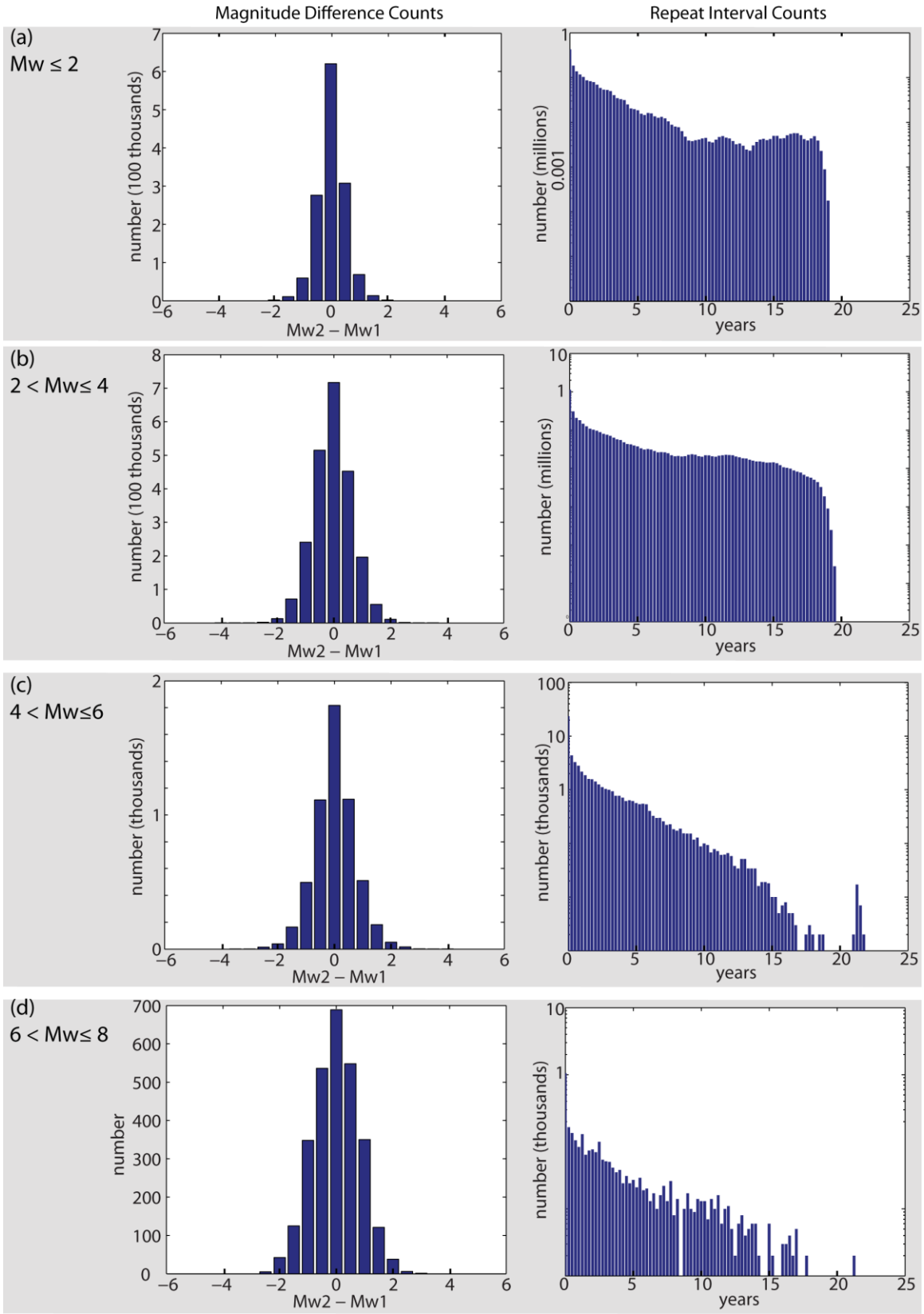
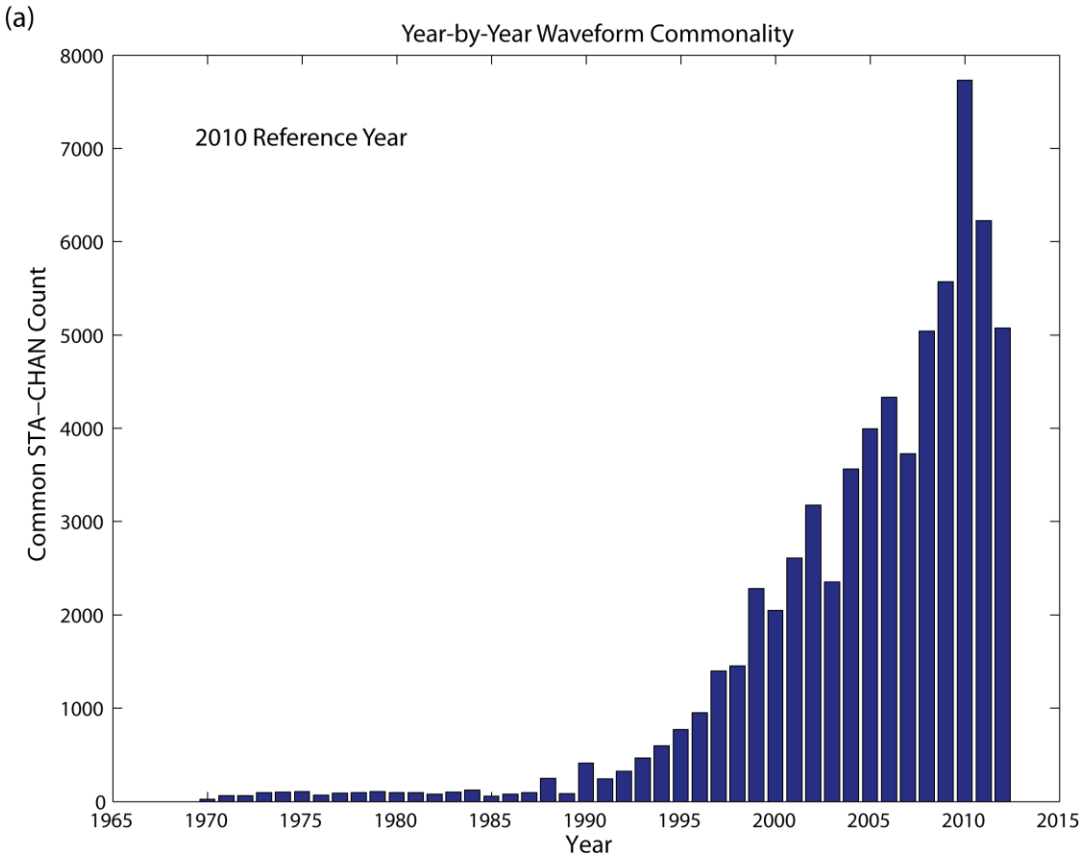
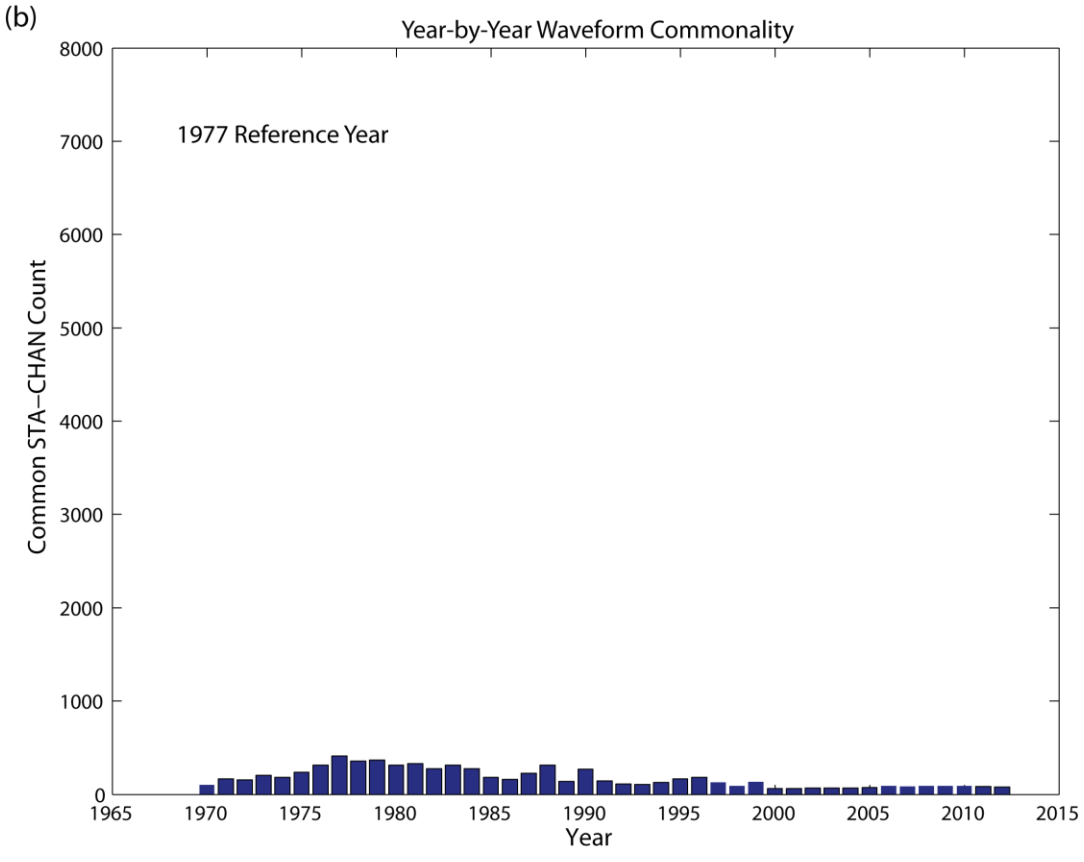


Figure 7 shows the magnitude differences (left) and the distribution of time separations (right) for correlated event pairs in our results. The data are divided

into four bins based on the average magnitude of each event pair. Panel (a) shows results for average Mw ≤ 2. Panel (b) shows results for 2 < Mw (avg) ≤ 4. Panel(c) shows results for 4 < Mw (avg) ≤ 6, and panel (d) shows results for 6 < Mw (avg) ≤ 8.

835


(a)
Year-by-Year Waveform Commonality
2010 Reference Year
Common STA–CHAN Count
Year
0
1000
2000
3000
4000
5000
6000
7000
8000
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015


(b)
Year-by-Year Waveform Commonality
1977 Reference Year
Common STA–CHAN Count
Year
0
1000
2000
3000
4000
5000
6000
7000
8000
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015

836

Figure 8 is a comparison of STA-CHAN waveform commonality on a year-by-year basis. Panel (a) uses 2010 as the reference year. It was produced by computing the intersection of the sets of waveform STA-CHAN each year with the set of waveform STA-CHAN in 2010. Note that until 1990 there are only tens of channels in common, but the number rises quite rapidly after 1990. Panel (b) was produced using 1977 as the reference year. It is scaled the same as (a) to show the relative size of the two data sets. Panel (b) also shows that only a few tens of STA-CHAN are common between the two data sets.

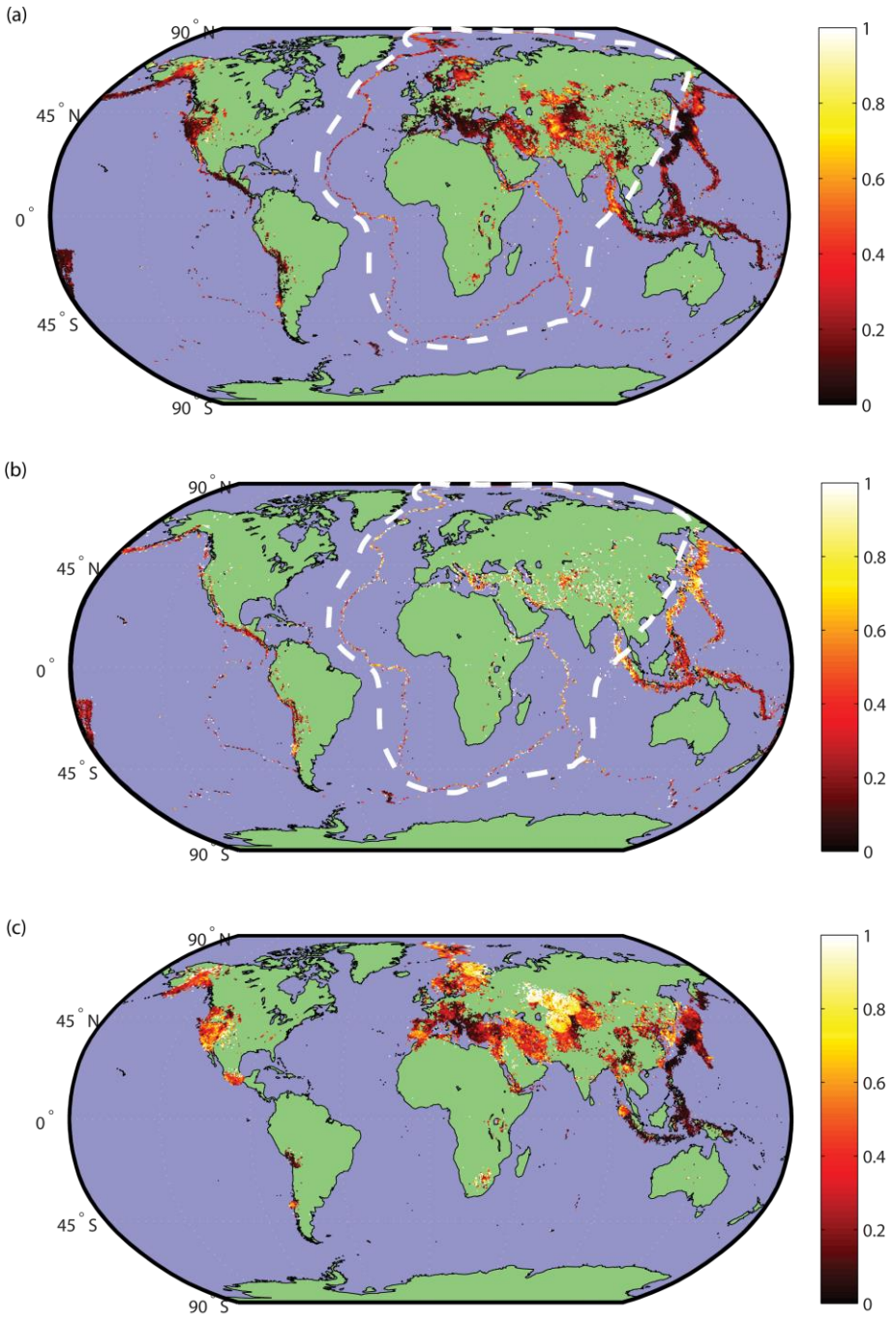
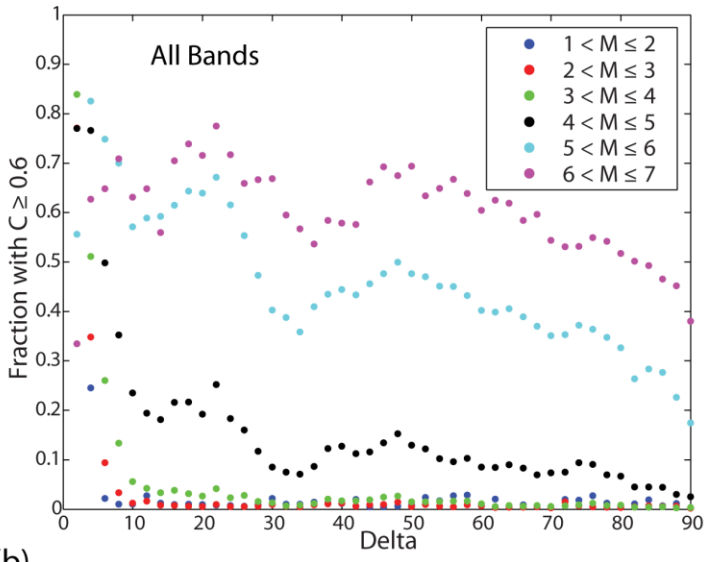
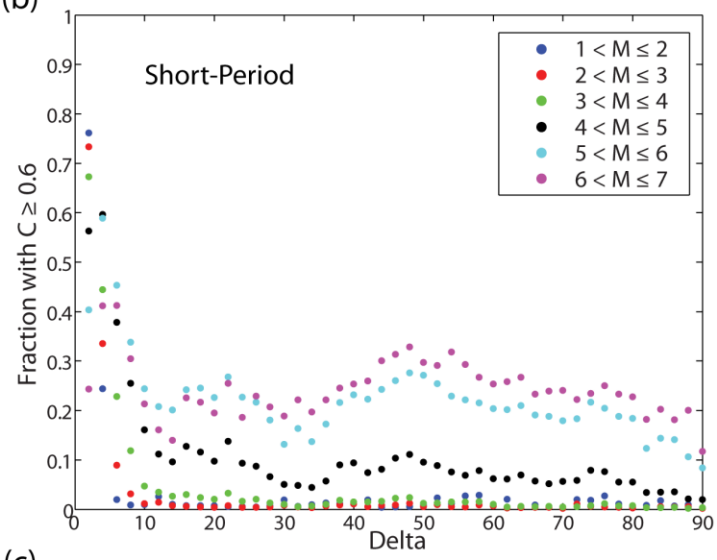


Figure 9 shows the geographic distribution of correlated events color-coded according to correlation fraction. The correlation fraction is defined as the number of events in a cell that correlate with at least one other event divided by the total number of bulletin events in the cell for the time period in which there are waveforms in the cell. Panel (a) shows the correlation fraction for all events. The dashed white line outlines the largest region in which our waveform holdings are at least 80% complete. Panel (b) shows the correlation fraction computed using only events >= Mw 5. Panel (c) shows the fraction of events for which we have waveforms for stations within five degrees of the epicenters.

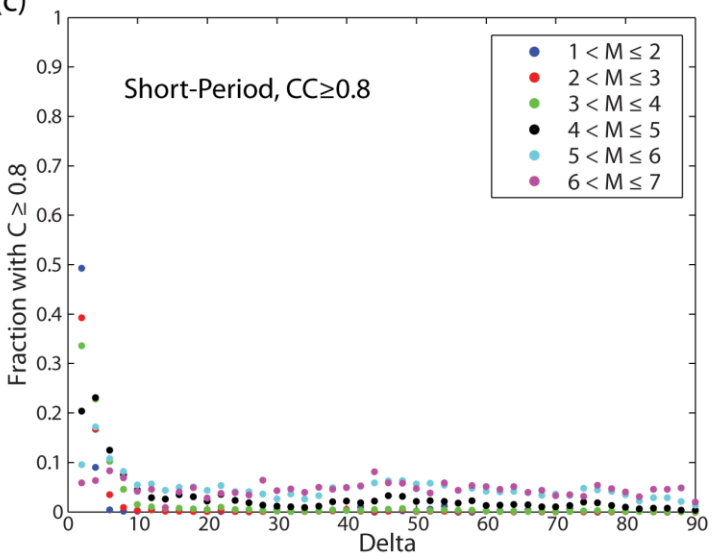
(a)


All Bands
1 < M ≤ 2
2 < M ≤ 3
3 < M ≤ 4
4 < M ≤ 5
5 < M ≤ 6
6 < M ≤ 7
Fraction with C ≥ 0.6
Delta

(b)


Short-Period
1 < M ≤ 2
2 < M ≤ 3
3 < M ≤ 4
4 < M ≤ 5
5 < M ≤ 6
6 < M ≤ 7
Fraction with C ≥ 0.6
Delta

(c)


Short-Period, CC≥0.8
1 < M ≤ 2
2 < M ≤ 3
3 < M ≤ 4
4 < M ≤ 5
5 < M ≤ 6
6 < M ≤ 7
Fraction with C ≥ 0.8
Delta

856

Figure 10 shows the fraction of correlated seismicity as a function of source-station separation in different magnitude ranges. Panel (a) shows the fraction of Catalog Events with Correlations in All Bands for 6 Mw Ranges. Panel (b) shows the fraction of Catalog Events with Correlations in Short-period Bands for 6 Mw Ranges. Panel (c) shows the fraction of Catalog Events with High Correlations (C >= 0.8) in Short-period Bands for 6 Mw Ranges.

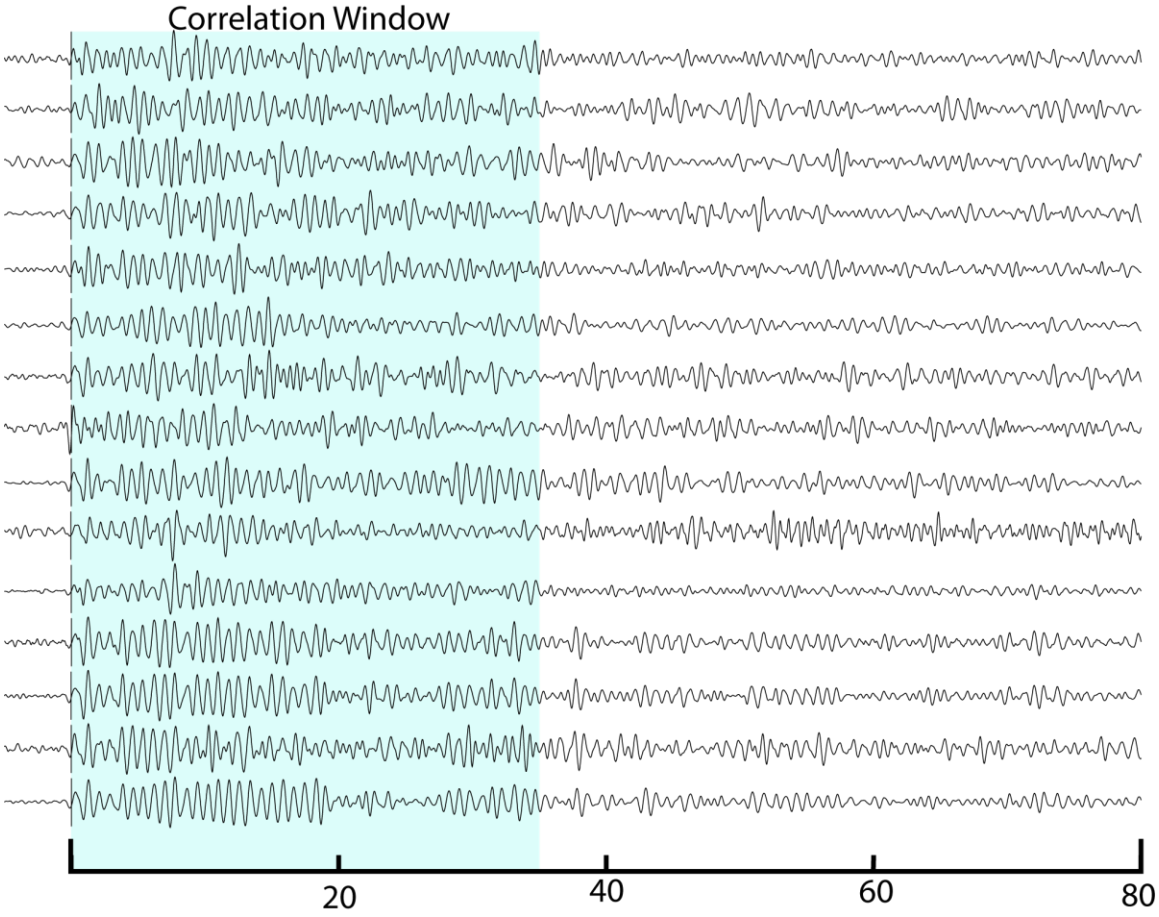


Figure 11 shows 80s-long seismograms recorded at KK01 for 15 events found to be mutually correlated in the 1-2 Hz band at the >= 0.6 level (average correlation was 0.75). The source-receiver separation was between 48 and 50 degrees, and the average correlation window length was ~35s.

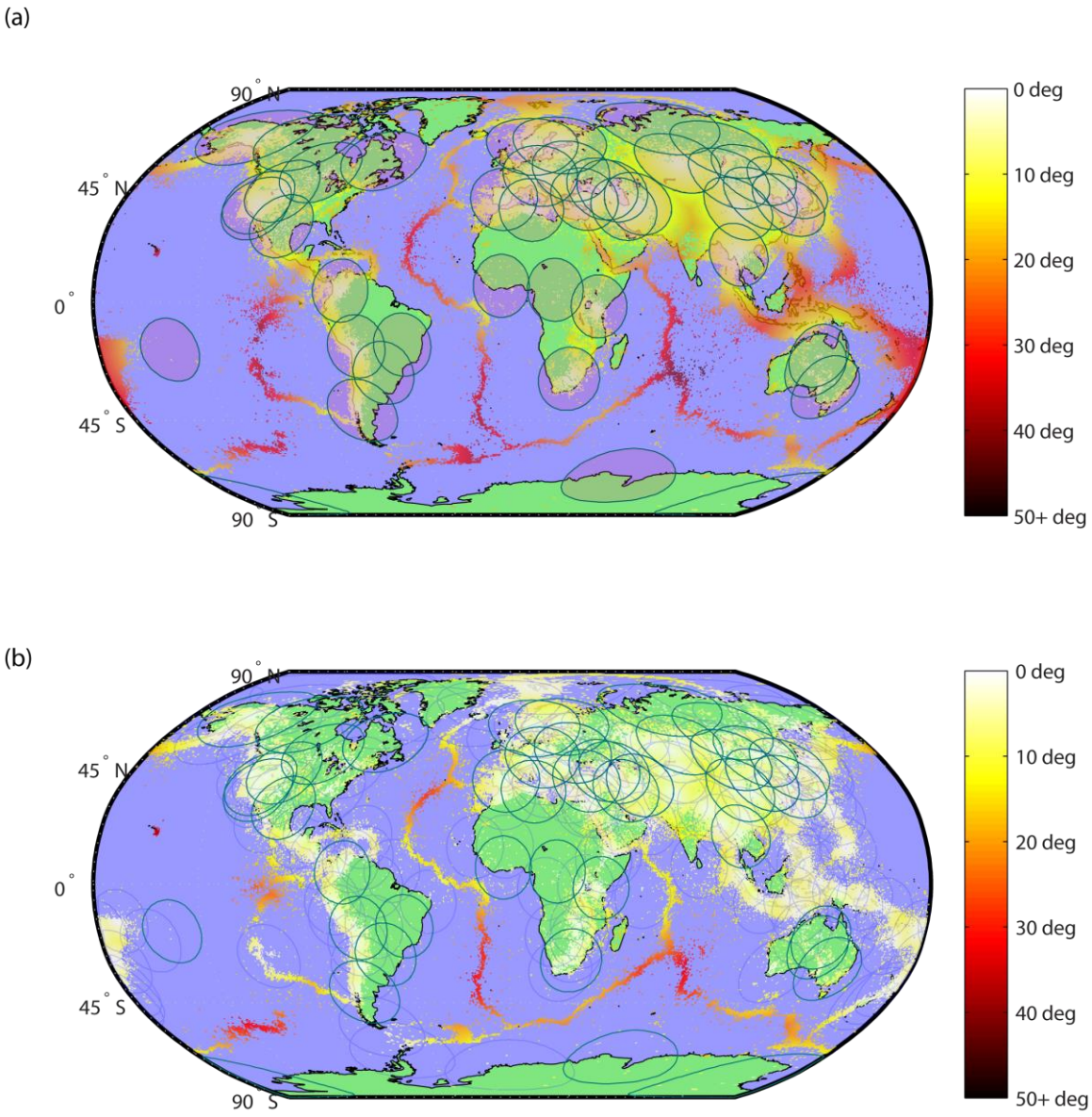


Figure 12 is a map of seismicity from the LLNL combined bulletin color coded according to distance from the nearest IMS station or array. Colors range from black for distances greater than 50 degrees to white for distance = 0. The small-circles are of 12 degree radius and are centered on IMS stations or arrays. Based on previous results, this is the effective bounding distance at which a substantial fraction of correlated waveforms may be observed in high frequency bands. Panel (a) shows the IMS primary stations and panel (b) shows the results for all IMS stations and arrays.